PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G06K 9/00 — A1

(11) International Publication Number: **WO 99/42950**

(43) International Publication Date: 26 August 1999 (26.08.99)

(21) International Application Number: PCT/US99/03288

(22) International Filing Date: 23 February 1999 (23.02.99)

(30) Priority Data:
09/027,685 23 February 1998 (23.02.98) US

(71) Applicant: ARCH DEVELOPMENT CORPORATION [US/US]; Room 405, 5640 South Ellis Avenue, Chicago, IL 60637 (US).

(72) Inventors: XU, Xin-Wei; 1042 Park Crest Drive, Darien, IL 60561 (US). DOI, Kunio; 6415 Lane Court, Willowbrook, IL 60521 (US).

(74) Agents: KUESTERS, Eckhard, H. et al.; Oblon, Spivak, McClelland, Maier & Neustadt, P.C., Suite 400, 1755 Jefferson Davis Highway, Arlington, VA 22202 (US).

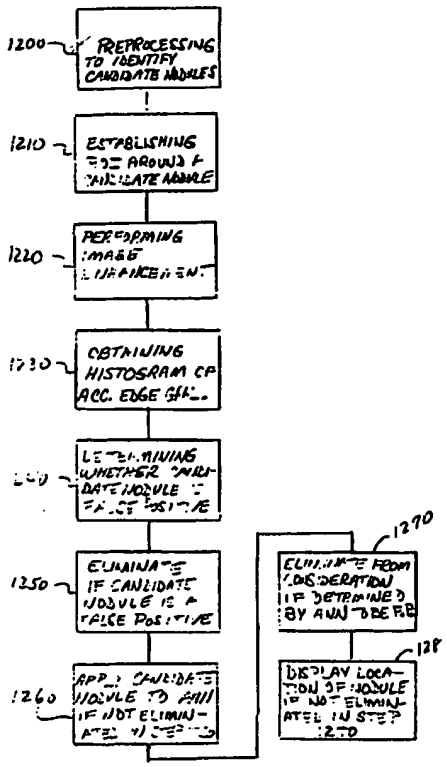
(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: LUNG NODULE DETECTION USING EDGE GRADIENT HISTOGRAMS

(57) Abstract

An automated method and a computer storage medium storing instructions for executing the method, for analysis of image features in lung nodule detection in a digital chest radiographic image, including preprocessing the image to identify candidate nodules in the image (1200); establishing a region of interest (ROI) including a candidate nodule within the ROI (1210); performing image enhancement of the candidate nodule within the ROI (1220); obtaining a histogram of accumulated edge gradients as a function of radial angles within the ROI after performing the image enhancement (1230); and determining whether the candidate nodule is a false positive based on the obtained histogram (1240).



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

LUNG NODULE DETECTION USING EDGE GRADIENT HISTOGRAMS

CROSS-REFERENCE TO RELATED APPLICATIONS

The present is related to automated techniques for automated detection of abnormalities in digital images, for example as disclosed in one or more of U.S. Patents 4,839,807; 4,841,555; 4,851,984; 4,875,165; 4,907,156; 4,918,534; 5,072,384; 5,133,020; 5,150,292; 5,224,177; 5,289,374; 5,319,549; 5,343,390; 5,359,513; 5,452,367; 5,463,548; 5,491,627; 5,537,485; 5,598,481; 5,622,171; 5,638,458; 5,657,362; 5,666,434; 5,673,332; 5,668,888; as well as U.S. applications 08/158,388; 08,173,935; 08/220,917; 08/398,307; 5,732,697; 5,740,268; 5,790,690; 5,832,103; 08/523,210; 08/536,149; 08/536,450; 08/562,087; 08/757,611; 08/758,438; 08/900,188; 08/900,189; 08/900,191; 08/900,192; 08/900,361; 08/900,362; 08/979,623; 08/979,639; 08/982,282; 09/028,518; and 09/027,468, each of which are incorporated herein by reference in their entirety. Of these patents and patent applications, U.S. patent 4,907,156 and U.S. patent application 08/562,087 are especially pertinent.

The present invention also relates to technologies referenced and described in the references identified in the appended APPENDIX and cross-referenced throughout the specification by reference to the number, in brackets, of the respective reference listed in the APPENDIX, the entire contents of which are also incorporated herein by reference. Various of these publications may correspond to various of the cross-referenced patents and patent applications.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

The present invention was made in part with U.S. Government support under grant CA 62625 from the USPHS. The U.S. Government has certain rights in the invention.

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates to an automated method and apparatus for lung nodule detection in chest radiographs, and more particularly to an automated method and apparatus in which lung nodules are detected in chest radiographs with a reduction in false positive detection.

Discussion of the Background

X-ray chest radiography is the most commonly used radiological imaging modality for detection of solitary subtle lung nodules in patients because of the low radiation does, low cost, and reliability. Solitary lung nodules in chest images are one of the important signs of primary lung cancer, which is the leading cause of cancer death in men and women in the United States.[1,2] However, it is well known that radiologists may fail to detect lung nodules in as many as 20 to 30% of actually positive cases viewed retrospectively.[3-6] In an effort to help radiologists to improve their diagnostic accuracy, at the Department of Radiology at the University of Chicago (UC), a computer-aided diagnosis (CAD) scheme for automated detection of lung nodules in chest images, as disclosed in above-noted U.S. patent application 08/562,087, has been developed.[7,8] A radiologist may use the computer output as a second opinion in making his/her diagnosis.[9,10]

The UC CAD scheme begins with a difference image technique, as disclosed in U.S. patent 4,907,156 [11], in which a nodule-suppressed image is subtracted from a nodule-enhanced image to produce a so-called difference image for reduction of normal background structures in the chest image. Nodule candidates in the chest image are selected by multiple gray-level thresholding of the difference image.[7,12] The derived nodule candidates are then classified into six groups according to the levels used by the multiple gray-level thresholding. The adaptive rule-based image feature analysis method is applied to nodule candidates in each group for removal of the corresponding false positives in each group. Finally, an artificial neural network (ANN) is trained to identify the candidates remaining after the rule-based tests.[7,8] For the UC database, which consisted of 200 PA chest images, including 100 normals and 100 abnormals (with 122 confirmed nodules), the prior UC CAD scheme achieves a performance of 70% sensitivity with 1.7 false positives per chest image.

It has been found that the majority of false-positive detections resulting from the prior UC CAD scheme are related to rib-rib or rib-vessel crossings, and that some others are due to shadows of soft tissues such as breast, heart, and diaphragm. The prior UC method for elimination of false-positive detections in the CAD scheme is based on gray levels and morphologic features obtained by the region-growing technique. These image features are derived from both the difference image and the original chest image.[7,8]

SUMMARY OF THE INVENTION

Accordingly, an object of this invention is to improve upon the prior UC CAD scheme, by reducing the number of false positives in the automated detection of lung nodules in chest radiographs.

Another object of this invention is to provide a novel method and apparatus for the automated detection of lung nodules in chest radiographs, utilizing a set of new features derived from the analysis of the histogram of accumulated radial edge gradients in an effort to improve the performance of the prior UC CAD scheme further by reducing the number of false positives.

These and other objects are achieved according to the present invention by providing a novel automated method and apparatus for the detection of lung nodules in chest radiographs, by incorporating new features derived from analysis of the histogram of radial edge gradients on nodule candidates. The present invention includes the recognition that approximately 80% of false positives are due to rib-rib or rib-vessel crossings, and also to interactions between ribs and soft tissues, such as breast, cardiac, or diaphragm shadows. A 64 X 64-pixel region of interest (ROI) centered at the candidate location is selected first. The contrast of the ROI is improved by a two-dimensional background subtraction. A nodule shape matched filter is used for enhancement of the nodular pattern located in the central area of the ROI. A histogram of accumulated edge gradients as a function of the radial angles is obtained. Analysis of the histogram results in seven features, including the maximum, minimum, width, and standard deviation of the histogram in a selected range of radial angles. The histogram from a "true" nodule ROI tends to have a narrow, prominent peak with a large maximum value near the radial axis. However, the rib structures generally broaden the corresponding histogram, thus resulting in a large width and a high minimum value. Features derived from the histogram are used for identifying some subtle and difficult false positives that can not be

eliminated by the prior UC CAD scheme. A rule-based test, combining all seven features, is applied with the elimination of 138 (40%) of 340 false positives without any loss of nodules. Then, an artificial neural network (ANN) is applied to remove an additional 8% of the remaining false positives with a reduction of 5% of true nodules, whereby the performance of prior UC CAD scheme is improved.

BRIEF DESCRIPTION OF THE DRAWINGS

A more complete appreciation of the invention and many of the attendant advantages thereof will be readily obtained as the same becomes better understood by reference to the following detailed description when considered in connection with the accompanying drawings, wherein:

Figures 1(a), 1(b) and 1(c) are illustrations of a region of interest (ROI) in a lung image with a nodule overlapped with a posterior rib, respectively illustrating (a) an original ROI; (b) the same ROI with background subtracted; (c) the same ROI after matched filtering;

Figure 1(d) is a graph of the radial edge gradient histogram of the filtered ROI illustrated in Figure 1(c);

Figure 2 is an illustration of geometric parameters defining the radial angle (β) for the edge gradient G at a pixel (x_i,y_i);

Figures 3(a), 3(b) and 3(c) are illustrations of a false positive ROI due to the interaction between the rib and breast shadow, respectively illustrating (a) an original ROI, (b) a ROI with background subtracted, and (c) a ROI after matched filtering;

Figure 3(d) is a graph illustrating the radial edge gradient histogram of the filtered ROI in Figure 3(c);

Figures 4(a), 4(b) and 4(c) are illustrations of a false positive ROI due to crossings of rib-clavicle or rib-rib, respectively illustrating (a) an original ROI, (b) a ROI with background subtracted, and (c) a ROI after matched filtering;

Figure 4(d) is a graph illustrating the radial edge gradient histogram of the filtered ROI in Figure 4(c);

Figures 5(a), 5(b) and 5(c) are illustrations of a false positive ROI due to rib-vessel crossing, respectively illustrating (a) an original ROI, (b) a ROI with background subtracted, and (c) a ROI after matched filtering;

Figure 5(d) is a graph illustrating the radial edge gradient histogram of the filtered ROI in Figure 5(c);

Figure 6 is an illustration of the relationship between the histogram width and the difference between the maximum and the minimum value, for nodules and false positives;

Figure 7 is an illustration of the relationship between the maximum and minimum values for nodules and false positives;

Figure 8 is an illustration of the relationship between the standard deviation and the partial standard deviation for nodules and false positives;

Figure 9 is an illustration of the relationship between the partial average value and the partial standard deviation for nodules and false positives; and

Figure 10 is an illustration of the relationship between the ratio of two maximum values near and outside the radial axis and the partial average value for nodules and false positives;

Figure 11 is a schematic illustration of a general purpose computer 100 programmed according to the teachings of the present invention; and

Figure 12 is a flowchart showing the steps performed according to the method of the invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The database employed in derivation of the present invention included 200 posterior-anterior (PA) chest radiographs, each 14" X 17". There were one hundred abnormal cases with a total of 122 nodules (confirmed by CT scans or radiographic follow-up) and one hundred normal cases (all verified by CT scans). The digitization procedure of the films as well as the size, contrast, and the subjective subtle rating of the 122 nodules are well documented elsewhere.[7,8] In the new study resulting in derivation of the present invention, the nodule candidates resulting from the prior UC CAD scheme were used. It is noted that, for this database, the prior scheme had a performance of 70% sensitivity with 1.7 false positives per chest image. Therefore, a total of 426 candidates (86 of 122 confirmed nodules and 340 false positives) produced by the scheme were used.

A 64 X 64-pixel region of interest (ROI) with a pixel size of approximately 0.7 mm centered at the candidate location was used in the new study. Figure 1(a) shows the ROI of

a nodule that overlaps a posterior rib. It is seen that the contrast of the nodule pattern is relatively low. To increase the contrast of the nodule, for each pixel in the ROI, the background subtraction was applied by using the average pixel value. This average value was obtained by taking the average of the pixel values along the column and also the row which were through that pixel. The increase in contrast of the nodule pattern by means of the background subtraction technique is demonstrated in Figure 1(b). A 9 mm nodule shape matched filter was then used to enhance the nodule pattern in the ROI further, as shown in Figure 1(c). The operation of this filtering was performed in the frequency domain by using Fast Fourier Transform (FFT).[13] It is clear in Figure 1(c) that the signal-to-noise ratio of the nodule pattern at the center of the ROI was enhanced considerably.

After the preprocessing by which the nodular pattern in the ROI was enhanced, the edge gradient, G, at each pixel was calculated by use of a 3 X 3 Sobel filter.[14] By accumulating edge gradients obtained from all pixels in the ROI, at each bin of different radial angles, the histogram of accumulated radial edge gradients was obtained. The radial angle (β) of the edge gradient G at a pixel is defined as the angle between the direction for the maximum edge gradient and the radial axis, which passes through the origin (x_0, y_0) of the ROI and that pixel (x_1, y_1), as illustrated in Figure 2. In Figure 2, the location (x_0, y_0) is the center (or origin) of the ROI, and (x_1, y_1) is the point where the edge gradient is considered. The radial angle is zero when the direction of the maximum edge gradient agrees with the radial axis, and has a range from -180° to 180°. The radial angle is positive when the gradient G rotates from the radial axis, around the point (x_1, y_1), counterclockwise and negative when it rotates from the radial axis clockwise. The radial edge gradient histogram describes the orientations of edge gradients relative to the radial axis. If an ideal round shape nodule with a Gaussian distribution would be located at the center of the ROI, the corresponding histogram will have a sharp peak at the radial angle of zero, because the direction of the edge gradient at every pixel inside the nodule pattern aligns with the radial axis. The histogram of radial edge gradients for the nodule ROI shown in Figure 1 is plotted in Figure 1(d). It is seen that the prominent peak in the histogram is shifted slightly to a radial angle of approximately 30°. This is probably because the shape of the nodule is close to an ellipse rather than a circle.

According to the present invention, a total of seven features were derived by analysis of the histogram of accumulated radial edge gradients for the purpose of elimination of some "difficult" false positives which could not be removed by the prior CAD scheme. These features are: (1) the maximum histogram value (or the maximum value) near the radial axis, which is defined here for a range of the radial angles from -90° to +90°; (2) the minimum histogram value (or the minimum value) near the radial axis; (3) the partial average value, which is the average of the histogram values between the minimum and 65% of the maximum; (4) the standard deviation of histogram values near the radial axis; (5) the standard deviation of histogram values used for the calculation of the partial average value (this standard deviation is referred to as the partial standard deviation); (6) the width of the histogram (in terms of degrees) including both sides from zero degree of the radial angle, at the histogram value corresponding to 35% of the difference between the maximum and the minimum value; (7) the ratio of the maximum value near the radial axis to the maximum value in the two outside ranges, corresponding to the range of the radial angles from -180° to -90° and from 90° to 180°. Hereinafter, these features are discussed in detail in terms of their properties as well as their usefulness in distinguishing between lung nodules and false positives.

For the nodule ROI shown in Figure 1, there is a prominent peak in the histogram near the radial axis (from -90° to +90°), as illustrated in Figure 1(d). Therefore, the maximum value and the standard deviation for the nodule tend to be large. On the other hand, if the structures are located around the peripheral areas of the ROI, their edge gradients tend to contribute in two outside ranges of the radial angles, namely, from -180° to -90° and from 90° to 180°, because the edge gradients of peripheral structures generally deviate largely from the radial direction. It is evident in Figure 1(c) that the nodule shape matched filter enhanced the nodule considerably, which is located at the center of the ROI. Therefore, this resulted in most edge gradients of the ROI (where the major pattern is a nodule at the center) to be distributed near the radial axis. Consequently, the ratio of the maximum value near the radial axis to the corresponding maximum value in the two outside ranges becomes very large.

Figure 3 shows a false positive ROI which resulted from the interaction between the rib and breast (the soft tissue) shadow. Clearly, the pattern around the center of this ROI

did not resemble the shape of the nodular. Consequently, the matched filter used did not enhance the structure around the center of the ROI effectively, as shown in Figure 3(c). The radial edge gradient histogram obtained from the filtered ROI is illustrated in Figure 3(d). It is seen that the maximum value near the radial axis is small. On the other hand, the peripheral structures contributed to a large amount of edge gradients in the outside range of the radial angles from -180° to -90°. Therefore, the ratio of the maximum value near the radial axis to that in the two outside ranges became small.

An example of a false positive ROI due to bone crossings (rib-rib or rib-clavicle) is shown in Figure 4. It is clear that the rib-clavicle crossing in the central area of the ROI was enhanced by the nodule shape matched filter (Figure 4(c)) because the appearance of the rib-clavicle or rib-rib crossing is generally similar to that of a nodule. This enhancement resulted in a large maximum value near the radial axis, as is shown in Figure 4(d). However, the bone structures which extend from the center of the ROI to its four corners were also enhanced. The edge gradients of these linear bony patterns could have a wide range of radial angles and thus tend to broaden the histogram width, as shown in Figure 4(d). In addition, it should be noted that, in Figure 4(d), the minimum histogram value near the radial axis became large as the width of the histogram increased.

Finally, a false positive ROI resulting from rib-vessel crossings is demonstrated in Figure 5. The histogram of radial edge gradients included multiple peaks in the range near the radial axis, probably because the patterns of rib-vessel crossings in the central area of the ROI were not uniformly enhanced by the matched filter used, as shown in Figure 5(c). The standard deviation of the histogram near the radial axis was small. However, the rib and vessel structures on the left side of the ROI contributed a significant amount of edge gradients in the range of radial angles from -180° to -90°, as shown in Figure 5(d).

Figure 6 shows the relationship between the histogram width and the difference between the maximum and the minimum value. It is apparent in Figure 6 that some false positives (mainly due to rib-rib or rib-clavicle crossings) have a large histogram width. About 34 out of 340 false positives (10% of the total number of false positives used in this study) had the histogram width larger than 180 degrees. Also, it is useful to note in Figure 6 that the difference between the maximum and the minimum value tends to be large for nodules compared with that for some false positives. This is because, as shown in Figure

7, a nodule was likely to have a histogram with a large maximum value as well as a small minimum value. This result is also consistent with the observation of Figure 1(d), which presents the histogram of a nodule ROI. Thus if the width of the histogram including both sides from zero degrees of the radial angle, at a histogram value equal to 35 % of the difference between the maximum and minimum histogram values exceeds 180°, it is considered that the detected nodule is a false positive and is eliminated from consideration.

Similarly, from Figure 7, it is seen that no nodules but a substantial number of false positives, about 21% of the false positives (71 of 340), have a minimum value larger than 0.05. Since no nodules have a minimum histogram value larger than 0.05, it is possible to eliminate from consideration as a nodule all such false positives having a minimum histogram value larger than 0.05.

As also seen from Figure 7, no nodules and a number of false positives have a maximum histogram value less than approximately 0.08. Therefore, it is further possible to eliminate from consideration as a nodule those false positives having a maximum histogram value less than 0.08. At a cutoff value of maximum histogram value equal to 0.08, about 12 of 340 false positives will be removed without loss of nodules. This is about 3.5% reduction of remaining false positives

The standard deviation of a histogram indicates the magnitude of its variation. The histogram of a nodule (Figure 1(d)) usually has a prominent peak near the radial axis and thus results in a large standard deviation. However, histograms of some false positives such as rib-vessel crossings were relatively flat near the radial axis that yielded small standard deviations (see Figure 5(d)). In Figure 8, the standard deviation and the partial standard deviation of histograms are compared for nodules and false positives. About 6% (21 of 340) of false positives and no nodules had a standard deviation below 0.03. It is therefor possible to eliminate from consideration as nodules those false positives having a standard deviation below 0.03. As shown in Figure 8, it is seen that the standard deviation and the partial standard deviation have some correlation.

The partial average value and the partial standard deviation of the histogram were obtained by ignoring the contribution from those radial edge gradients around the prominent peak. It is noted that the edge gradients around the prominent peak were most

likely resulted from the nodule shape pattern (i.e., nodules or rib-rib crossings) at the center of the ROI. Notice also that the appearance of some rib-rib or rib-bone (such as rib-clavicle) crossings at the central area of ROI was very similar to that of true nodules. However, these two features, i.e., the partial average value and the partial standard deviation, were related to edge gradients due to structures away from the central area of the ROI. Clearly, the rib-rib crossings such as those seen in Figure 4, tended to have large partial average values due to the fact that their minimum values near the radial axis were large. A scatter plot for comparison of the partial average value and the partial standard deviation for nodules and false positives is shown in Figure 9. As evident in Figure 9, no nodules and about 17% of the total number of false positives (58 of 340) have a partial standard deviation less than 0.005. It is therefor possible to eliminate from consideration as nodules those false positives having a partial standard deviation less than 0.005. It is noted that false positives tend to have a small partial standard deviation or a large partial average value.

From Figure 9 it is also seen that no nodules and a number of false positives exhibit a partial average value greater than 0.0725. It is therefore possible to eliminate from consideration as a nodule those false positives have a partial average value greater than 0.0725. For the partial average value in Fig 9, at the cutoff value of 0.0725, about 19 of 340 false positives will be removed without loss of nodules. This is about 5.5% reduction of remaining false positives.

To evaluate the effect of the matched filter used on the central pattern of the ROI, the ratio of the maximum value near the radial axis (from -90° to 90°) to that in the two outside ranges (from -180° to -90° and from 90° to 180°) was calculated. For a nodule located at the central area of the ROI, its pattern could be enhanced more effectively by the matched filter used, thus resulting in a prominent peak in the histogram near the radial axis and a corresponding large ratio. However, for false positives such as soft tissue or some rib-vessel crossings, their structures did not closely match the filter used and had a relatively flat histogram near the radial axis compared with that in the two outside ranges. The ratios tend to be smaller for some false positives than for nodules, as shown in Figure 10. In fact, no nodules and about 10% of the total number of false positives (34 of 340) have a ratio less than 0.5. Therefore, the ratio is a useful feature, indicating how effective

the matched filter is on the central pattern of the ROI, and providing another basis for eliminating from consideration as a nodule all those false positives having a ratio of less than 0.5. It is further noted that the ratios of many rib-rib or rib-clavicle crossings were also large. Ratios of these false positives were overlapped with those of nodules, as shown in Figure 10. This is because these crossings were also enhanced effectively by the matched filter used.

For the combined sequential rule-based test, a candidate was detected as a nodule by the computer if its features had satisfied all of the test rules one by one. A further reduction of false positives was possible by application of an artificial neural network (ANN) for the remaining 202 false positives.[7,15] The ANN included seven input units corresponding to the seven features derived from the radial edge gradient histogram, five hidden units, and one output unit. A candidate was identified as a "true" nodule or false positive if the output value from the output unit was close to 1 or 0, respectively. A jackknife test [7,15] was employed by training the ANN with one half of the remaining false positives and testing with the other half. It was found that approximately 8% of the remaining false positives were eliminated by the ANN, but with a reduction of 5% of true nodules. Therefore, with the application of the technique of the present invention, it is possible to make a further improvement in the performance of the prior CAD scheme for automated detection of lung nodules on digital chest images.

This invention may be conveniently implemented using a conventional general purpose digital computer or micro-processor programmed according to the teachings of the present specification, as will be apparent to those skilled in the computer art. Appropriate software coding can readily be prepared by skilled programmers based on the teachings of the present disclosure, as will be apparent to those skilled in the software art.

The present invention includes a computer program product which is a storage medium including instructions which can be used to program a computer to perform a process of the invention. The storage medium can include, but is not limited to, any type of disk including floppy disks, optical discs, CD-ROMs, and magneto-optical disks, ROMs, RAMs, EPROMs, EEPROMs, magnetic or optical cards, or any type of media suitable for storing electronic instructions.

Figure 11 is a schematic illustration of a general purpose computer 100 programmed

according to the teachings of the present invention. The general purpose computer 100 includes a computer housing 102 having a motherboard 104 which contains a CPU 106 and memory 108. The computer 100 also includes plural input devices, e.g., a keyboard 122 and mouse 124, and a display card 110 for controlling monitor 120. In addition, the computer system 100 further includes a floppy disk drive 114 and other removable media devices (e.g., tape, and removable magneto-optical media (not shown)), a hard disk 112, or other fixed, high density media drives, connected using an appropriate device bus, e.g., a SCSI bus or an Enhanced IDE bus. Also connected to the same device bus or another device bus, the computer 100 may additionally include a compact disc reader/writer 118 or a compact disc jukebox (not shown).

Stored on any one of the above described storage medium (computer readable media), the present invention includes programming for controlling both the hardware of the computer 100 and for enabling the computer 100 to interact with a human user. Such programming may include, but is not limited to, software for implementation of device drivers, operating systems, and user applications. Such computer readable media further includes programming or software instructions to direct the general purpose computer 100 to perform tasks in accordance with the present invention.

The programming of general purpose computer 100 may include a software module for digitizing and storing PA radiographs obtained from an image acquisition device. Alternatively, it should be understood that the present invention can also be implemented to process digital data derived from a PA radiograph elsewhere. Otherwise, the computer 100 includes software modules implementing the method of the invention shown in Figure 12, including preprocessing the image to identify candidate nodules in the image (step 1200); establishing a region of interest (ROI) including a candidate nodule identified in the preprocessing step (step 1210); performing image enhancement of the candidate nodule within the ROI (step 1220); obtaining a histogram of accumulated edge gradients as a function of radial angles within the ROI after performing step 1220 (step 1230); determining whether the candidate nodule is a false positive based on the obtained histogram (step 1240); and eliminating from consideration false positives identified in the determining step (step 1250).

Step 1240 includes determining at least one evaluation factor, including at least one

of (1) a maximum histogram value, (2) a minimum histogram value, (3) a partial average value of the histogram, (4) a standard deviation of the histogram values near the radial axis, (5) a partial standard deviation of histogram values, (6) a width of the histogram including both sides from zero degrees of the radial angle, at a predetermined histogram value, and (7) a ratio of a maximum histogram value near the radial axis to a maximum histogram value in two predetermined outside ranges of the radial axis. Step 1250 further includes eliminating said candidate nodule from consideration if the determined at least one evaluation factor for said candidate nodule is greater than, or less than, a predetermined threshold, in dependence on the respective evaluation factor.

Generally, in step 1240 each evaluation factor will be evaluated sequentially, thereby to maximize the number of false positives which can be eliminated from further consideration as a nodule. However, the present invention is not limited to the evaluation of each named evaluation fact, and fewer than all the evaluation factors can be evaluated if economy of processing time or processing capacity is imperative.

As shown in Figure 12, the method of the invention further includes step 1260, application of an artificial neural network (ANN) in an effort to identify remaining false positives. In step 1260, a candidate is identified as a "true" nodule or false positive if the output value from the ANN output unit is close to 1 or 0, respectively. In step 1270, those candidates identified as a false positive are also removed from consideration as a nodule, Then, in step 1280, the locations of remaining candidate nodules, i.e., those original candidate nodules remaining after elimination of false positives, are displayed for viewing by a radiologist.

The invention may also be implemented by the preparation of application specific integrated circuits or by interconnecting an appropriate network of conventional component circuits, as will be readily apparent to those skilled in the art.

Recapitulating, the present invention is based on the recognition that the features derived from the analysis of the radial edge gradient histogram are effective in eliminating some subtle and difficult false positives which could not be eliminated by the prior CAD scheme.[7,8] By combining all of the cutoff rules related to these features sequentially, for example, the histogram width larger than 180°, the minimum value larger than 0.05, the standard deviation less than 0.03, the partial standard deviation less than 0.005, and the

ratio less than 0.5, a total of 138 (40%) false positives are eliminated without any loss of nodules. For the combined sequential rule-based test, a candidate is detected as a nodule by the computer if its features have satisfied all of the test rules one by one. A further reduction of false positives is possible by application of an artificial neural network (ANN) for the remaining 202 false positives.[7,15] The ANN included seven input units corresponding to the seven features derived from the radial edge gradient histogram, five hidden units, and one output unit. A candidate is identified as a "true" nodule or false positive if the output value from the output unit is close to 1 or 0, respectively. A jackknife test [7,15] is employed by training the ANN with one half of the remaining false positives and testing with the other half. Approximately 8% of the remaining false positives are eliminated by the ANN, but with a reduction of 5% of true nodules. Therefore, with the application of the technique of the present invention, it is possible to make a further improvement in the performance of the prior CAD scheme for automated detection of lung nodules on digital chest images.

Obviously, numerous modifications and variations of the present invention are possible in light of the above teachings. It is therefore to be understood that within the scope of the appended claims, the invention may be practiced otherwise than as specifically described herein.

V. REFERENCES

1 *Cancer facts and figures - 1996*, Atlanta: American Cancer Society, 1996.

2 *Annual cancer statistics review: Including cancer trends 1950-1985*, National Cancer Institute, 1988.

3 J. Yerushalmy, "Reliability of chest radiography in the diagnosis of pulmonary lesions," Am. J. Surg. **89**, 231-240 (1955).

4 L.W. Guiss and P. Kuenstler, "A retrospective view of survey photofluorograms of persons with lung cancer," Cancer **13**, 91-95 (1960).

5 J.V. Forrest and P.J. Friedman, "Radiologic errors in patients with lung cancer," West J. Med. **134**, 485-490 (1981).

6 J.R. Muhm, W.E. Miller, R.S. Fontana, D.R. Sanderson, and M.A. Uhlenhopp, "Lung cancer detected during a screening program using four-month radiographs," Radiology **148**, 609-615 (1983).

7 X.W. Xu, K. Doi, T. Kobayashi, H. MacMahon, and M.L. Giger, "Development of an improved CAD scheme for automated detection of lung nodules in digital chest images," Med. Phys. **24**, 1395-1403 (1997).

8 X.W. Xu, H. MacMahon, M.L. Giger, and K. Doi, "Adaptive feature analysis of false positives for computerized detection of lung nodules in digital chest images" SPIE Medical Imaging 1997, Image Processing Vol. **3034** 428-436 (1997).

9 K. Doi, M.L. Giger, H. MacMahon, K.R. Hoffmann, S. Katsuragawa, R.M. Nishikawa, Y. Yoshimura, S. Sanada, X. Chen, C.E. Metz, C.J. Vyborny, R.A. Schmidt, S.M. Montner, T. Matsumoto. and K.G. Chua, "Computer-aided diagnosis: present and future, "*A new horizon on medical physics and biomedical engineering* H. Abe, K. Atsumi, T. linuma, M. Saito, and M. Inoue, editors; Elsevier Science Publishers B.V. 1991.

10 K. Doi, M. L. Giger, R. M. Nishikawa, K. R. Hoffmann, H. MacMahon, R. A. Schmidt, and K. G. Chua, "Digital radiography: A useful clinical tool for computer-aided diagnosis by quantitative analysis of radiographic images," Acta Radiologica **34**, 426-439 (1993).

11 M.L. Giger, K. Doi, and H. MacMahon, "Image feature analysis and computer-aided diagnosis in digital radiography. 3. Automated detection of nodules in peripheral lung fields," Med. Phys. **15**, 158-166 (1988).

12 M.L. Giger, K. Doi, H. MacMahon, C.E. Metz, and F.F. Yin, "Pulmonary nodules: Computer-aided detection in digital chest image," RadioGraphics **10**, 41-51 (1990).

13 E.O. Brigham, *The Fast Fourier Transform and its Applications,* Prentice-Hall, Inc. Englewood Cliffs, New Jersey, 1988.

14 I. Pitas, *Digital Image processing Algorithtm,* Prentice Hall International (UK) Ltd, 1993.

15 Y.C. Wu, K. Doi, M.L. Giger, C.E. Metz, and W. Zhang, "Reduction of false positives in computerized detection of lung nodules in chest radiographs using artificial neural networks, discriminate analysis and a rule-based scheme," J. of Digital Imag 7, 196-207 (1994).

Claims:

1. An automated method for analysis of image features in lung nodule detection in a chest radiographic image represented by digital data, comprising:

preprocessing said image to identify candidate nodules in said image;

establishing a region of interest (ROI) including a candidate nodule identified in said preprocessing step;

performing image enhancement of said candidate nodule within said ROI;

obtaining a histogram of accumulated edge gradients as a function of radial angles within said ROI after performing said image enhancement; and

determining whether said candidate nodule is a false positive based on the obtained histogram;

eliminating from consideration false positives identified in said determining step.

2. The method of Claim 1, wherein:

said determining step comprises determining at least one evaluation factor, including at least one of (1) a maximum histogram value, (2) a minimum histogram value, (3) a partial average value of the histogram, (4) a standard deviation of the histogram values near the radial axis, (5) a partial standard deviation of histogram values, (6) a width of the histogram including both sides from zero degrees of the radial angle, at a predetermined histogram value, and (7) a ratio of a maximum histogram value near the radial axis to a maximum histogram value in two predetermined outside ranges of the radial axis; and

said eliminating step comprises eliminating said candidate nodule from consideration as a nodule if the determined at least one evaluation factor for said candidate nodule has a value on a predetermined side of a predetermined threshold value.

3. The method of Claim 2, wherein:

said determining step comprises determining the maximum histogram value in a predetermined range near the radial axis as an evaluation factor; and

said eliminating step comprises eliminating said candidate nodule from consideration as a nodule if the determined maximum histogram value is less than a predetermined value.

4. The method of Claim 2, wherein:

said determining step comprises determining a minimum histogram value near the

radial axis as an evaluation factor; and

said eliminating step comprises eliminating said candidate nodule from consideration as a nodule if the determined minimum histogram value near the radial axis exceeds a predetermined threshold.

5. The method of Claim 2, wherein:

said determining step comprises determining a partial average value of the histogram in a portion of the ROI away from a central area of the ROI as an evaluation factor; and

said eliminating step comprises eliminating said candidate nodule from consideration as a nodule if the determined partial average value exceeds a predetermined value.

6. The method of Claim 2, wherein:

said determining step comprises determining a standard deviation of the histogram values near the radial axis of the histogram as an evaluation factor; and

said eliminating step comprises eliminating said candidate nodule from consideration as a nodule if the determined standard deviation is below a predetermined value.

7. The method of Claim 2, wherein:

said determining step comprises determining a partial standard deviation of histogram values in a portion of the ROI away from a central area of the ROI as an evaluation factor; and

said eliminating step comprises eliminating said candidate nodule from consideration as a nodule if the determined partial standard deviation is below a predetermined value.

8. The method of Claim 2, wherein:

said determining step comprises determining the width of the histogram, including both sides from zero degrees of the radial angle, at a predetermined histogram value as an evaluation factor; and

said eliminating step comprises eliminating said candidate nodule from consideration as a nodule if the determined width of the histogram exceeds a predetermined value.

9. The method of Claim 2, wherein:

said determining step comprises determining the ratio of a maximum histogram value near the radial axis to a maximum histogram value in two predetermined outside ranges of the radial axis as an evaluation factor; and

said eliminating step comprises eliminating said candidate nodule from consideration

as a nodule if the determined ratio is less than a predetermined value.

10. The method of Claim 2, further comprising:

applying said candidate nodule to an artificial neural network (ANN) and identifying said candidate as a nodule or false positive if an output value from an ANN output unit is close to 1 or 0, respectively.

11. The method of Claim 10, further comprising:

displaying a location of the candidate identified as a nodule.

12. The method of Claim 1, wherein said step of performing image enhancement comprises:

removing image background from said ROI.

13. The method of Claim 12 wherein said step of removing image background comprises:

determining the average of the pixel values along a column and along a row intersecting said candidate nodule; and

subtracting said average from pixel values of pixels within said ROI.

14. The method of Claim 1, wherein said step of performing image enhancement comprises:

performing shape matched filtering on pixel in said ROI including said candidate nodule.

15. The method of Claim 14, wherein said step of performing shape matched filtering comprises:

performing said shape matched filtering in the frequency domain using a fast Fourier Transform (FFT).

16. The method of Claim 1. wherein said step of obtaining a histogram comprises:

calculating an edge gradient at each pixel using a 3 x 3 Sobel filter; and

accumulating edge gradients from all pixels in the ROI, at each bin of different radial angles.

17. The method of Claim 2, wherein said determining step comprises determining each of evaluation factors (1) - (7).

18. The method of Claim 4, wherein:

said determining step comprises determining a partial standard deviation of

histogram values in a portion of the ROI away from a central area of the ROI as an evaluation factor; and

said eliminating step comprises eliminating said candidate nodule from consideration as a nodule if the determined partial standard deviation is below a predetermined value.

19. The method of Claim 4, wherein:

said determining step comprises determining the width of the histogram, including both sides from zero degrees of the radial angle, at a predetermined histogram value as an evaluation factor; and

said eliminating step comprises eliminating said candidate nodule from consideration as a nodule if the determined width of the histogram exceeds a predetermined value.

20. The method of Claim 4, wherein:

said determining step comprises determining the ratio of a maximum histogram value near the radial axis to a maximum histogram value in two predetermined outside ranges of the radial axis as an evaluation factor; and

said eliminating step comprises eliminating said candidate nodule from consideration as a nodule if the determined ratio is less than a predetermined value.

21. The method of Claim 4, wherein:

said determining step comprises determining a standard deviation of the histogram values near the radial axis of the histogram as an evaluation factor; and

said eliminating step comprises eliminating said candidate nodule from consideration as a nodule if the determined standard deviation is below a predetermined value.

22. The method of Claim 18, wherein:

said determining step comprises determining the width of the histogram, including both sides from zero degrees of the radial angle, at a predetermined histogram value as an evaluation factor; and

said eliminating step comprises eliminating said candidate nodule from consideration as a nodule if the determined width of the histogram exceeds a predetermined value.

23. The method of Claim 18, wherein:

said determining step comprises determining the ratio of a maximum histogram value near the radial axis to a maximum histogram value in two predetermined outside ranges of the radial axis as an evaluation factor; and

said eliminating step comprises eliminating said candidate nodule from consideration as a nodule if the determined ratio is less than a predetermined value.

24. The method of Claim 18, wherein:

said determining step comprises determining a standard deviation of the histogram values near the radial axis of the histogram as an evaluation factor; and

said eliminating step comprises eliminating said candidate nodule from consideration as a nodule if the determined standard deviation is below a predetermined value.

25. The method of Claim 22, wherein:

said determining step comprises determining the ratio of a maximum histogram value near the radial axis to a maximum histogram value in two predetermined outside ranges of the radial axis as an evaluation factor; and

said eliminating step comprises eliminating said candidate nodule from consideration as a nodule if the determined ratio is less than a predetermined value.

26. The method of Claim 22, wherein:

said determining step comprises determining a standard deviation of the histogram values near the radial axis of the histogram as an evaluation factor; and

said eliminating step comprises eliminating said candidate nodule from consideration as a nodule if the determined standard deviation is below a predetermined value.

27. The method of Claim 26, wherein:

said determining step comprises determining a standard deviation of the histogram values near the radial axis of the histogram as an evaluation factor; and

said eliminating step comprises eliminating said candidate nodule from consideration as a nodule if the determined standard deviation is below a predetermined value.

28. A computer readable medium storing computer instructions for identification of nodules in a chest radiographic image represented by digital data, by performing the steps of any one of Claims 1-27.

29. A system for implementing the method of any one of Claims 1-27.

Figure 1(a)

Figure 1(b)

Figure 1(c)

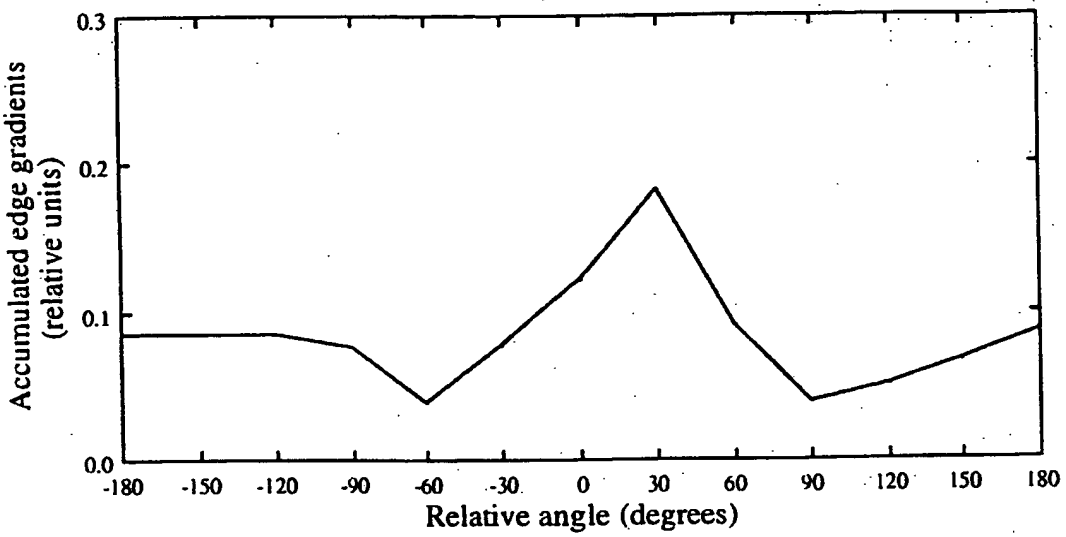


Figure 1(d)

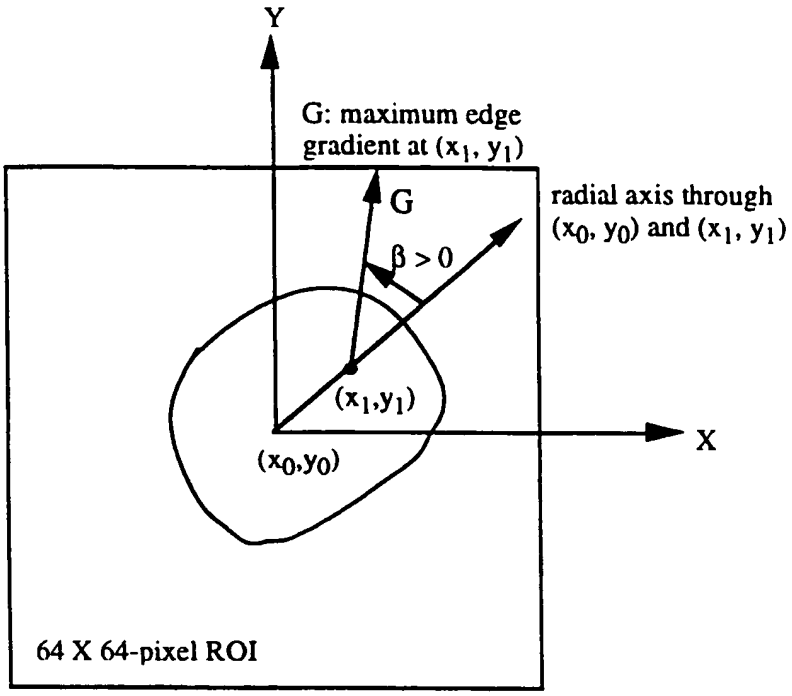

Y
G: maximum edge gradient at (x_1, y_1)
G
radial axis through (x_0, y_0) and (x_1, y_1)
$\beta > 0$
(x_1, y_1)
X
(x_0, y_0)
64 X 64-pixel ROI

Figure 2

Figure 3(a) Figure 3(b) Figure 3(c)

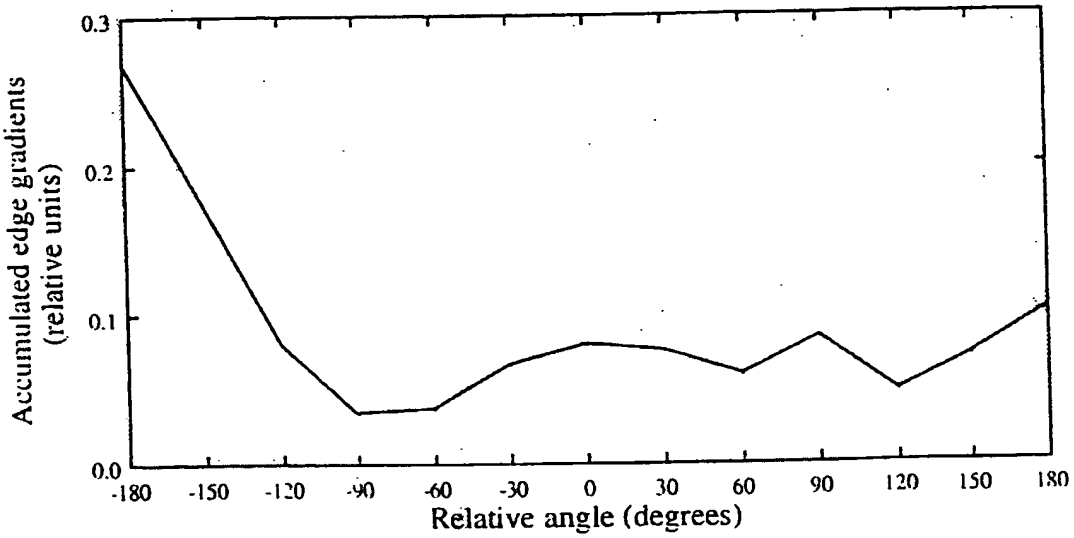


Figure 3(d)

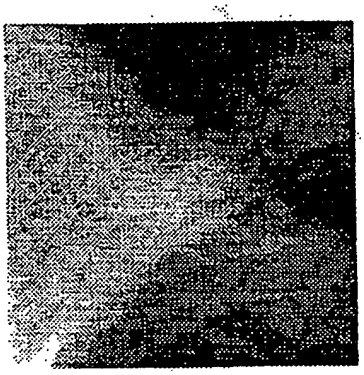


Figure 4(a)

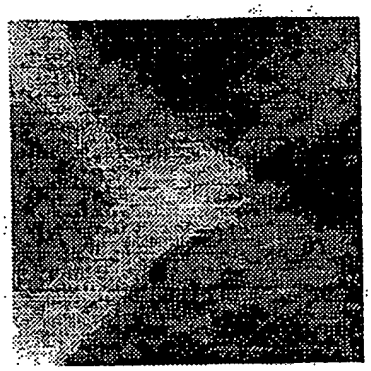


Figure 4(b)

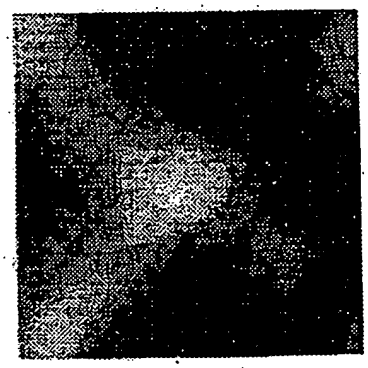


Figure 4(c)

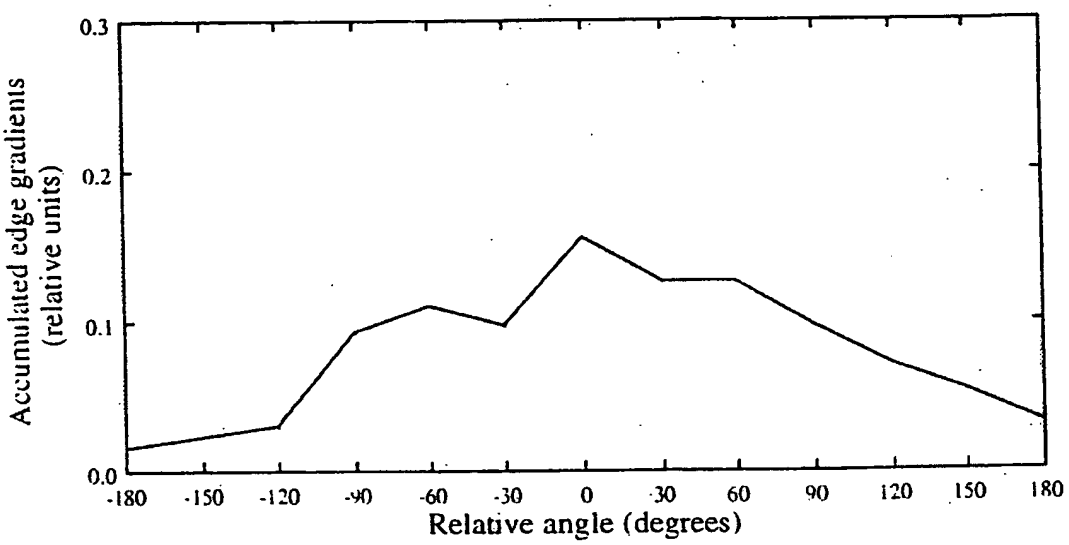


Figure 4(d)

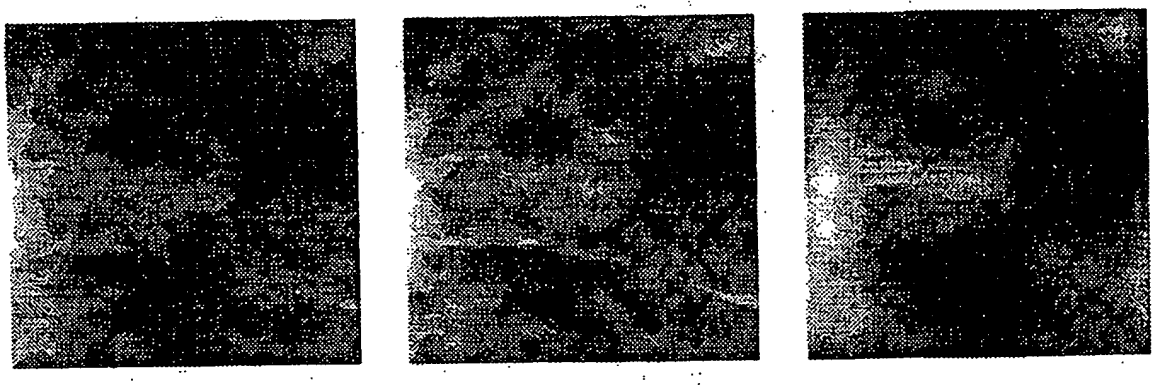


Figure 5(a) Figure 5(b) Figure 5(c)

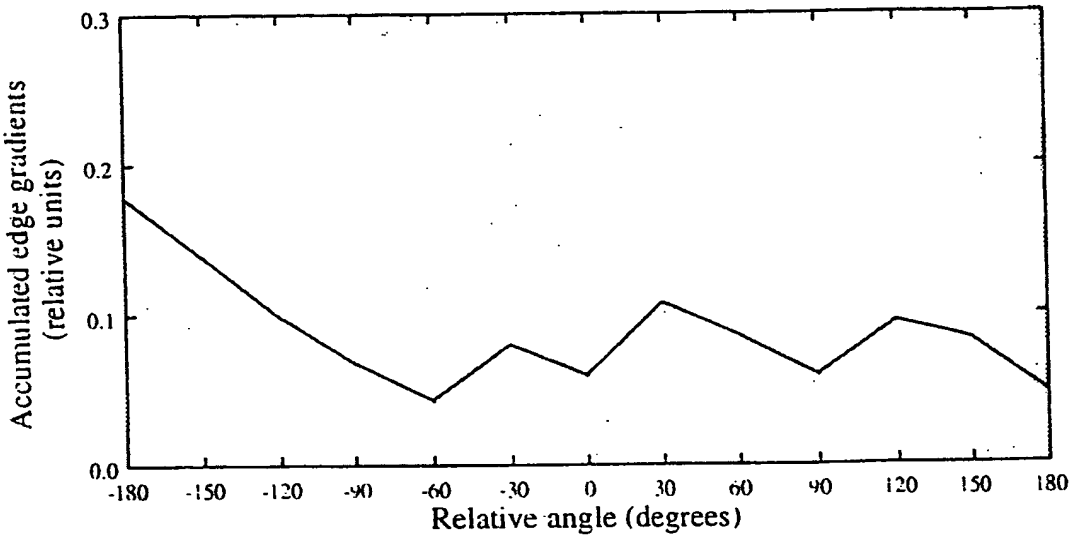


Figure 5(d)

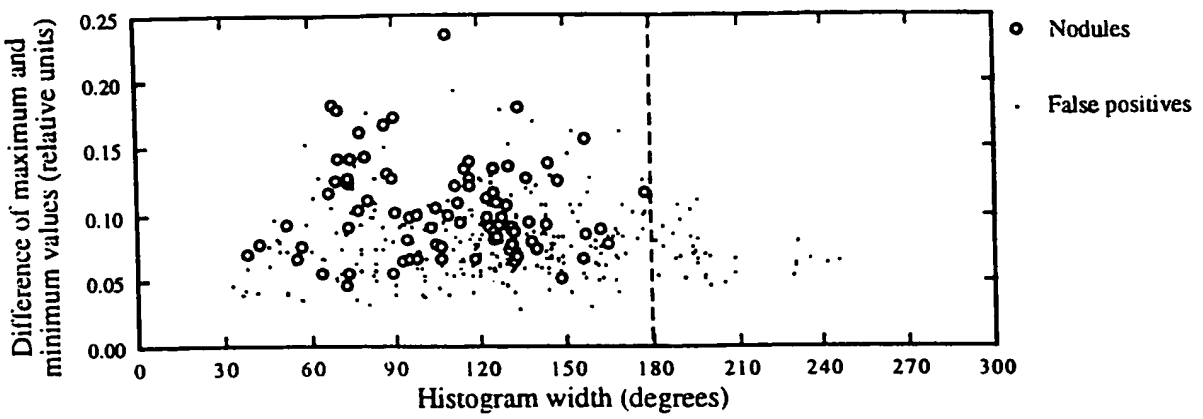


Figure 6

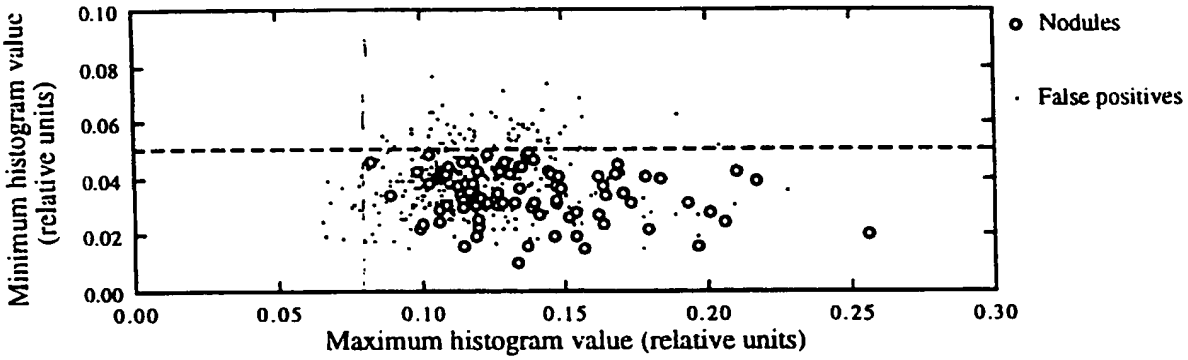


Figure 7

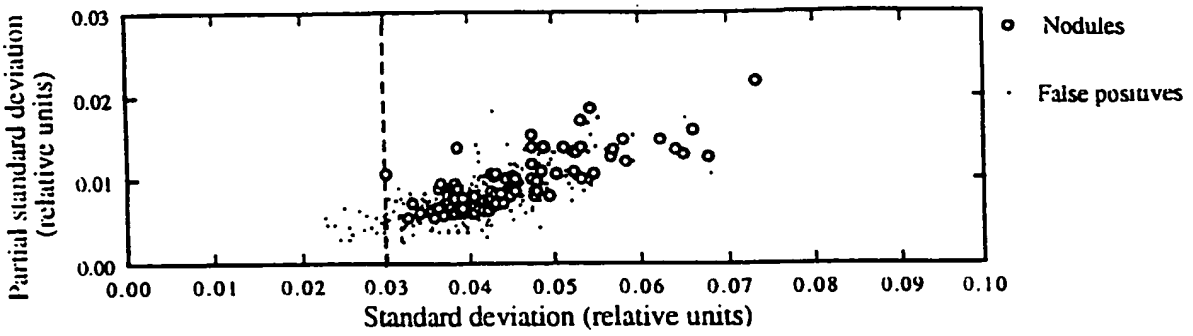


Figure 8

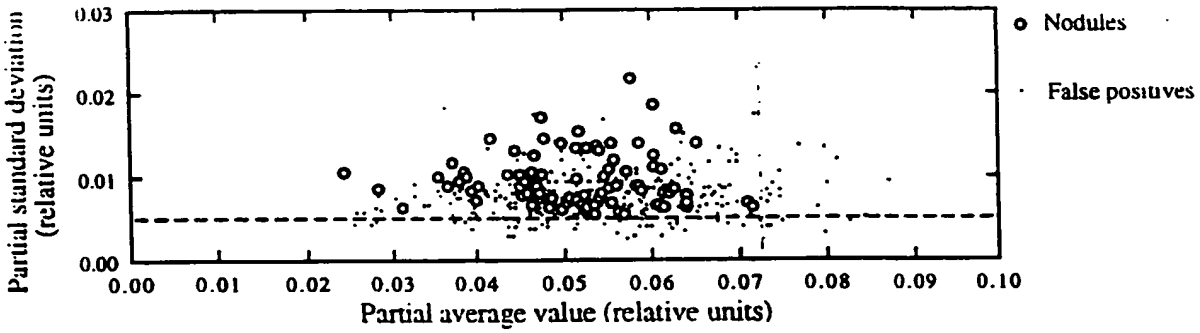


Figure 9

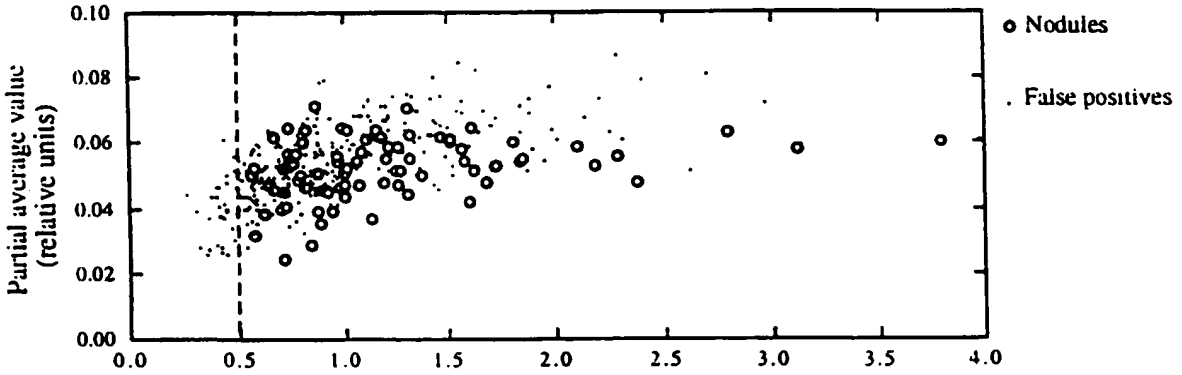


Figure 10

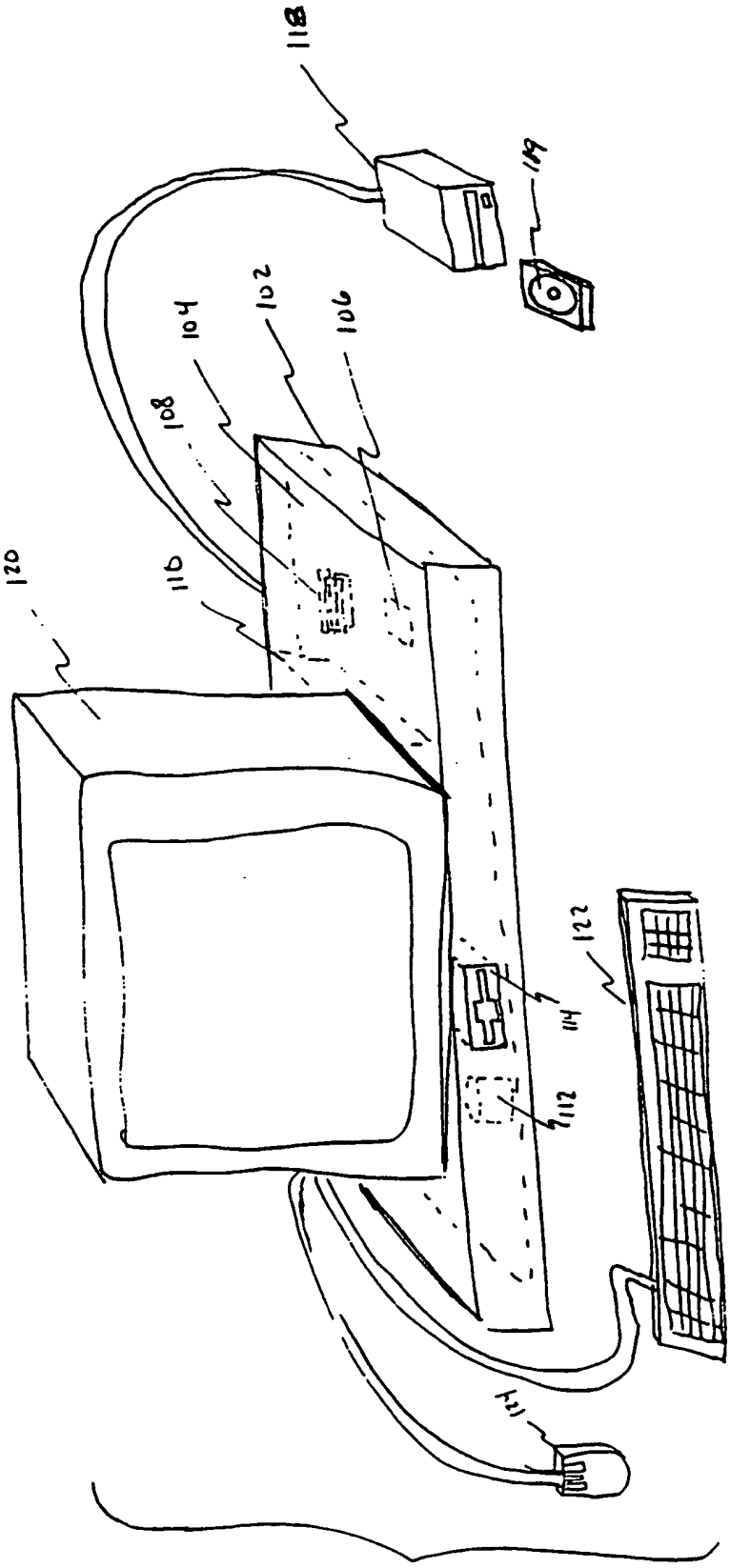


Figure 11

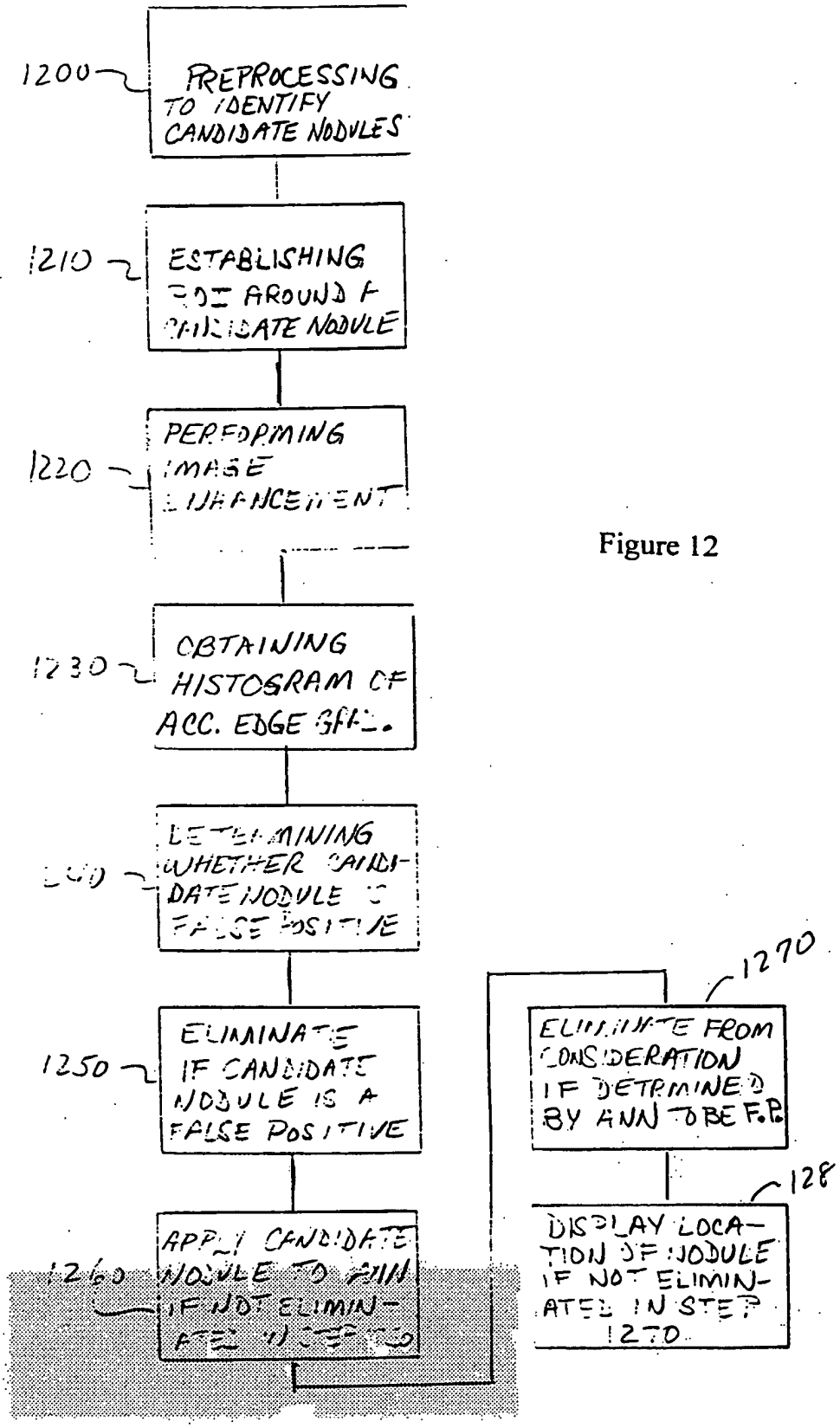

1200
PREPROCESSING TO IDENTIFY CANDIDATE NODULES
1210
ESTABLISHING
1220
PERFORMING
1230
OBTAINING HISTOGRAM OF
1240
1250
ELIMINATE IF CANDIDATE NODULE IS A FALSE POSITIVE
1260
1270
ELIMINATE FROM CONSIDERATION IF DETRMINED BY ANN TO BE F.P.
128
DISPLAY LOCATION OF NODULE IF NOT ELIMINATED IN STEP 1270

Figure 12

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/03288

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :G06K 9/00
US CL :382/132

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 382/132, 168, 199; 128/922

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS
search terms: edge-gradient histogram, lung, nodule, false positives

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5,289,374 A (DOI et al) 22 February 1994, entire reference.	1-2, 5-7, 12
---		---
Y		3-4, 8-11, 13-29
Y	US 5,491,627 A (ZHANG et al) 13 February 1996, abstract.	3-4, 8-11, 13-29
Y	US 4,907,156 A (DOI et al) 06 March 1990, col.5, line 67 to col.6, line 3.	14-15
Y	US 5,319,549 A (KATSURAGAWA et al) 07 June 1994, col.6, lines 38-41.	16

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
15 APRIL 1999	06 MAY 1999

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	JON CHANG Virginia Zagan
Facsimile No. (703) 305-3230	Telephone No. (703) 305-8439

Form PCT/ISA/210 (second sheet)(July 1992) ★